**Evaluation Project Article on HR Analytics and Attrition**

Human Resources forms the back bone of the industries world wide. Every company puts in lots of efforts, time and money to mould efficient and also loyal work force for their company. The companies also undertake various training programs for increasing the efficiency and also the effectiveness of their workforce. Even with all these efforts if the employee leaves the company for whatever reasons the company has to bear huge losses as the amount spent on training the previous employee is wasted and also now the company has to undertake more training for the new employee filling in the place. This can lead to huge losses and definitely analysis can help in deciding which employees are likely to undergo attrition and who might not. **This leads us to the problem statement of the data which is to predict based on various factors whether an employee might leave the company or continue with it.**

Before going further it is also imperative to discuss about HR analytics. Human resource analytics (HR analytics) is an area in the field of analytics that refers to applying analytic processes to the human resource department of an organization in the hope of improving employee performance and therefore getting a better return on investment. HR analytics does not just deal with gathering data on employee efficiency. Instead, **it aims to provide insight into each process by gathering data and then using it to make relevant decisions about how to improve these processes.** A part of this is also the prediction of attrition of the company. HR analytics helps in devising strategies for effective process building, motivation in work, compensation programs and a good work culture to retain the trained employees. If the company fails to retain their employees it has to go through once again the process of advertising, interviewing, hiring and training the new employees which leads to loss. Having regular attrition is thus a huge problem for the company. The company would also be devoid of higher collective knowledge and mastered skill sets which is gained through years of experience and this cannot be developed through just training new recruits. Also if the company has work related to frequent customer interaction, this would also be deemed as a grave loss as the customers are more comfortable to interact with the regular employees of the company. **So we can conclude attrition can lead to not only losses for moulding a new workforce but also the experience gained by the previous employees is also lost and this does not gel well with the customers as well.**

In the given data we observed the given data was in excel and hence we saved it in the local device and then used read\_excel to get the data into the Jupyter file. The data has 1470 rows and 35 columns in it, which are the features of the data and the second column is the label of the data which is Attrition. On subsequent analysis it was found the data has no nulls present in it and is relatively clean. The features of the data are Age, Business Travel, Daily Rate, Department, Distance From Home, Education, Education Field, Employee Count, Employee Number, Environment Satisfaction, Gender, Hourly Rate, Job involvement, Job level, Job role, Job satisfaction, Marital status, monthly income, Monthly rate, Number of companies Worked, Over18, Overtime, Percent Salary Hike, Performance rating, Relationship Satisfaction, Standard hours, Stock option level, Total working years, Training times last year, work life balance, years at company, years in current role, years since last promotion, years with current manager.

During the Exploratory Data Analysis, value counts of many features were taken along with the attrition to get an insight into the trend of attrition. It was observed for the one’s who travel frequently for Business travel the attrition was 24.9% which was the highest. So the one’s who travel frequently found it difficult to mange and this led to high attrition. For different Departments it was seen Sales and Human Resources had the highest attrition with 20% and 19% respectively. For the feature of Distance from home it was observed employees furthest from the workplace was more likely to discontinue the work. In the education feature it was observed employees from education level 1 had the highest 18% of attrition and then followed by level 3. A general trend observable was the one’s with the highest levels of Education chose to continue with their role in the company which could be because of the high levels of renumeration they might be receiving. For employees in different education field it was observed, the one’s in Human resources had the highest attrition followed by the technical degree and then marketing. For Environment Satisfaction it was noted the employees in level 1 was having the highest attrition and the ones at level 4 had the least. The people at level 4 was also the highest. Most people in the company had good levels of satisfaction. For the gender feature males had a little higher levels of attrition compared to females. For Job involvement the one’s having highest job involvement had the least attrition and the ones having less involvement had the highest attrition. For the 5 levels of Job level in the company the one’s in level 1 had very high levels of attrition. This is a very important factor for attrition. For the job roles the one’s as Sales representative had the highest attrition and the employees as Research Director had the least attrition. The one’s having Job satisfaction in level 1 had highest attrition. In the Marital Status feature the Single employees had the highest levels of attrition and the least attrition was for the Divorced. Employees who had worked in only 1 company showed the highest attrition followed by the one’s who worked in 9 companies. The employees who have worked in many companies show more like a settling trend and yet the ones who have worked in 9 companies also show attrition. For percent salary hike it was obvious the once having least hike showed the highest levels of attrition. The company had most employees in performance rating 3 and then in rating 4 and both classes showed the same percent of attrition. In relationship satisfaction it was noted the employees in level 1 of satisfaction had the highest percent (20.8%) of attrition showing the one’s who are not having a satisfactory relation had a high chance of leaving the company. For total working hours the one’s who had a higher working hours had a higher attrition. In the work life balance feature it was observed that the ones who had a better work life balance had less percent of attrition. In the years in company feature we observed the more years an employee spends in the company less is his chance of leaving the company although there is an element who worked for 40 years and then left the company but it is very possible the employee retired and did not leave the company for joining a new company. In the years in current role we noted the more years an employee works in a particular role less is his chance of leaving the company. The ones who had just taken up a particular role that is less than a year or aroud 2 years had the highest chance of leaving the company. In the years since last promotion feature we noted the employees who did not receive promotion in 1 year or less have left the company whereas some employees who had not been promoted for 15 years have not left the company. It is worth mentioning that the once who had joined the company recently expects better renumeration and promotion and the once who have been in the company for quite some time develop a sense of loyalty for the company and stay with it. But there are employees who have left the company after 15 years of not getting a promotion and the analytics should consider such employees as they have been with the company for long and have immense experience and expertise in their field. Also the employees who had new managers have also left the company more and the one’s who had the same manager have stayed with the company.

**In this exploratory analysis which was done by using value\_counts() we can conclude that employees who had to travel either for business or for office left the company for something more suited for them. It is necessary the company makes the employees are given jobs fit for their calibre and also they are felt included in their respective positions. The once who are at lower levels of work will undergo attrition and for this the company has to strategize ways of inclusiveness for such employees. The company should also check the job satisfaction of the employees and also their relationship status. A stable family person has lesser chance for attrition compared to singles. The employee who has worked in the company for many years can develop a sense of loyalty and hence can stay but he loss of such an employee is indeed a grave loss for the company as it can lead to loss of years of experience and expertise the employee possessed.**

The data was then split into categorical columns and numeric features and describe and distplot was taken which did imply outliers in some features like total working years and years at company. For analysing the data further Attrition which was the label was encoded using Label Encoder. **In the catplot which was taken later it was confirmed that in the marital status feature the married one’s have less attrition and also the divorced and the Singles had the highest attrition levels. For people in different ages it was noted the employees in the age group of 27 to 35 had a high density of attrition. Employees lesser than that age and more had less attrition density implying the employees lesser than 27 years of age continue in the company to develop experience and the ones more than 35 have found a stability with the company and hence continue with the company.** In the catplot for distance from home it is noticeable the density of the employees for level 1, 2, 3 and 9 and 10 are very high. But in the previous analysis the percent for level 29 was high which implies most employees prefer to have workplace close to their residence and hence the company had many employees in level 1 and eventhough there was attrition in them the percent was less whereas the employees in level 29 was less and a small number of people leaving showed a huge percent loss in level 29. In the next swarmplot it was noticeable the employees having level 1 of satisfaction and leaving the company were mostly singles and married employees. The company should consider the environment satisfaction of the employees to retain their workforce. In the pair plot we can see the employees leaving the company are mostly having low daily rates for their work. For distance from home and age we can see as the age increases the rate of attrition is low eventhough they have the same distance from home. An employee travelling the distance to the company develops a comfort in travel after some years and gets used to the daily travel and might not leave the country. **It is worth mentioning that most people leaving the company are young in age.** If the employee are adequately renumerated they might continue with the company. **Age has very high relation with the attrition. Young in age and having high monthly rates can still lead to attrition.**

All these are the results of the exploratory data analysis.

In the subsequent model building we used Ordinal encoder on the columns of EducationField, Gender, JobRole, Marital Status, Overtime and Department to give weights in order.We also dropped Employee Count, Employee Number and Over18 features of the data. In Business Travel we replaced the elements in order as Non-Travel, Travel Frequently and travel rarely. Thus the elements which showed more attrition were encoded higher values. The data was then checked for skewness and it was found to be skewed only for years since last promotion. We took the cube root of the data to remove the skewness of the data. The data was then checked for correlation and we found it have multicollinearity for feature of Monthly Income. The data was then split into features in x and label in y and was Standardised using Standard scalar. VIF was checked and Monthly income showed high VIF. After dropping Monthly Income the VIF was again checked and it was found to be in acceptable range. Since in the value counts the attrition values showed less number of Yes than No we used SMOTE to get the same value counts. Thus the pre processing steps were concluded and model building was started.

In the model building phase we used Random Forest Classifier to get a test model and also to obtain best random state for the data. Random Forest Classifier gave the best accuracy of 95.94% at random state 59. The data was then split into train and test at this random state. Different models were run on the tarin data by keeping the models and their names in a set. Of all the models extra tree classifier gave the best accuracy of 97% and the cross val score was also 0.002 which meant the data did not have much overfitting as well. The parameters of extra tree classifier were tuned and the best parameters obtained were criterion: entropy, max\_depth : 9, max\_features: log2 and min\_samples\_split : 2. However at these parameters the model gave an accuracy of 92% and hence with more trial and error we made a model with max\_depth 25 and min samples split 7 which gave accuracy of 96%

Later the model was saved.

We can safely conclude that attrition has high relation with Age of the employee, their Job involvement, Job satisfaction and Relationship Satisfaction as well. Features like Daily rate, Hourly rate are also of paramount importance. For the companies to retain their workforce it should focus on the age of the employee. The one’s in the age 27-35 should be given adequate opportunities to display their skills and talent and also be adequately renumerated. HR should focus on the involvement of the workforce and also their satisfaction to regulate the attrition. Attrition is not something that can be completely zero but with proper analysis its rate can be decreased which will help the company to avoid huge losses for training to get the same calibre in its workforce.